

### REMARKS

Applicants have studied the Office Action dated February 24, 2005. No new matter has been added. It is submitted that the application, is in condition for allowance. Applicants have amended Claims 1, 14, and 20. By virtue of this amendment, claims 1-20 are pending. Reconsideration and further examination of the pending claims in view of the above amendments and the following remarks is respectfully requested. In the Office Action, the Examiner:

- Rejected claims 1-3, 14-16, and 20 under 35 U.S.C. §103(a) as being unpatentable over Meyerzon et al. (U.S. Patent No. 6,638,314), in view of Lawrence et al. (U.S. Patent No. 6,289,342); and
- Rejected claims 4-6 and 17-19 under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (U.S. Patent No. 6,638,314), in view of Lawrence et al. (U.S. Patent No. 6,289,342) as applied to claim 1 and in further view of Hobbs (U.S. Patent No. 6,523,022).

### Telephonic Interview

Applicants wish to thank Examiner Burge for the telephone interview on Thursday, April 14, 2005. The Applicants' representatives Jon Gibbons and Tom Grzesik, participated in the telephone interview.

The discussion began with an overview of the present invention. The present invention provides a web crawling unit, a method, and a computer readable medium for assisting in the index of informational content at a given website by use of a search engine such as google™, yahoo™ and the like. One type of search program for indexing informational content at a website is known as a "web crawler." The web crawler creates an index of informational content for a given website for subsequent use by a search engine. The indexing by web crawlers is very useful for simple webpages where all the content is from one or two files. One problem often encountered by web crawlers is where information presented in a webpage is stored in secondary pages. The use of

ARC9-2000-0046-US1

8

09/607,370

secondary documents such as text, images, and other multimedia, make the management of webpage content much easier because each component of a webpage is broken down into pieces. Web crawlers do not load these secondary documents when indexing webpage content and this informational content is not properly indexed for a search engine.

Another problem encountered with using web crawlers to index informational content is where the webpage is dynamically assembled. Dynamically assembled information is information presented on a single webpage from more than one location. The dynamically assembled web content is not available from one location and must be gathered from several different storage locations before being presented as a single webpage to the user. This is especially true in websites where part of the informational content is retrieved from a database. The contents of the database must be gathered and assembled into a webpage format. Further, the use of client side scripts such as JavaScript and VBScript, do not make all the information available until the script is executed on the client side. Since web crawlers index informational content on server sites as opposed to client sites, the information content in client side script is not captured. The present invention solves the problem of web crawling dynamic data documents by temporarily rendering the information for the dynamic data document in memory in the manner the composer of the document intended it to be displayed. In this manner, the subsequent crawling indexes all the information on the website in the manner in which the creator of the web site intended the site to be viewed. This is in contrast to only discrete pieces of the informational content on the web site being indexed in the prior art. The present invention solves the problem with images containing textual data in part through use of optical character recognition on the images themselves to assist with web crawling.

To overcome the problems of using web crawlers to properly index dynamic websites containing secondary documents, with or without client-side scripts and the use of images with textual content, the present invention retrieves a web document at a given

address or URL. The contents of the web document are extracted for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Next, secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation are loaded. The in-memory webpage is rendered by the hub processing unit prior to analyzing and summarizing the in-memory webpage. The in-memory webpage representation is analyzed to produce a text map for the webpage document of the textual contents therein. The secondary documents include one or more images with textual content embedded therein. An optical character recognition engine is used on the images to extract textual content for adding to the textual map for the webpage document.

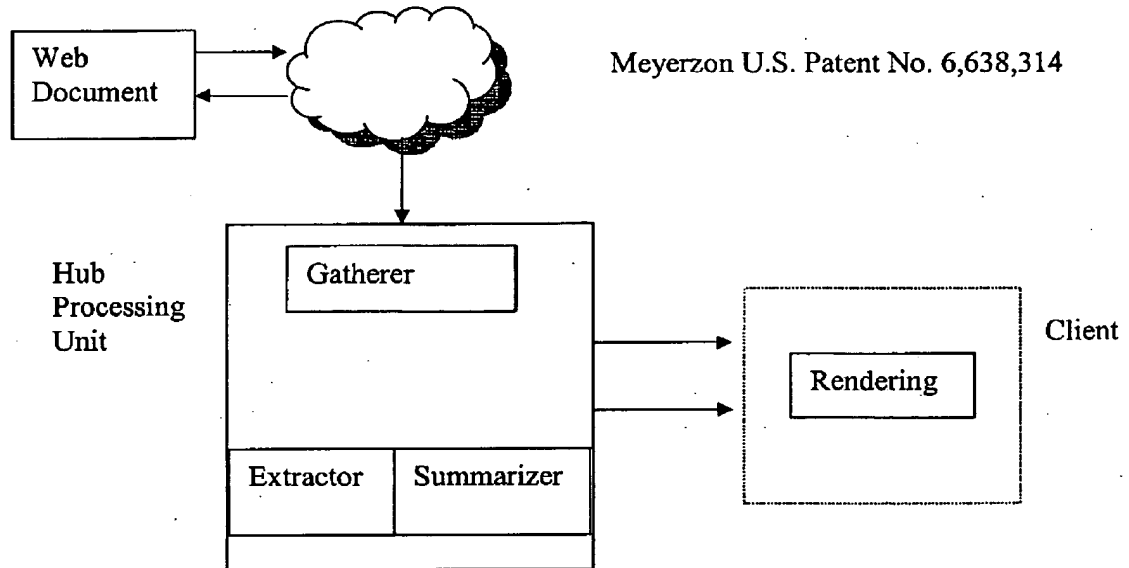
Discussed were the technical differences of the present invention, namely the web browser of the Meyerzon reference rendering a webpage on the client side and how this differs from the present invention of "rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser."

In order to more clearly point out this feature of a hub processing unit rendering an intermediate dynamically constructed in-memory webpage representation of the web document, which is formatted as if displayed for viewing on an end-user's web browser, the Applicants' representatives suggested the following language be added to the independent claims, i.e., claims 1, 14, and 20 as follows:

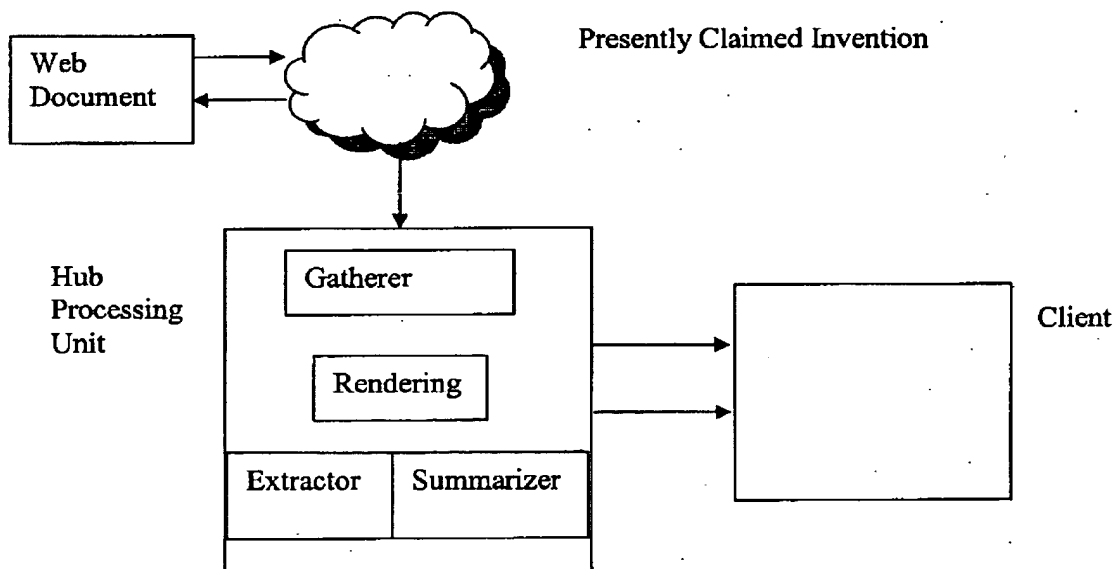
wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage

No new matter has been added. Support for this amended language is found in the application as originally filed and particularly at FIGS. 1 and 3a-3b and pages 9-14.

The following diagram illustrates how the Meyerzon renders the webpage at the client side only:



In contrast, the following diagram illustrates how the presently claimed invention renders an in-memory representation of the webpage:



As can be seen from the above diagrams, Meyerzon does not teach, suggest, or anticipate rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Meyerzon also does not teach, suggest, or anticipate loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein, wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage. Meyerzon renders the webpage at the client side and not before the hub processing unit analyzes and summarizes the in-memory webpage, as recited for independent claims 1, 14, and 20.

No agreement was reached on the claim language. The Examiner agreed to reconsider the claim language in this Amendment. The Applicants' representatives respectfully submit that all grounds for the rejection stated in the Office Action have been overcome. The Examiner is invited to call the Applicants' representative at the number listed at the end of this response, if she believes a telephone interview would advance prosecution.

Rejection under 35 U.S.C. §103(a) in view of Meyerzon et al. and Lawrence et al.

As noted above, the Examiner rejected claims 1-3, 14-16, and 20 under 35 U.S.C. §103(a) as being unpatentable over Meyerzon et al. (U.S. Patent No. 6,638,314), in view of Lawrence et al. (U.S. Patent No. 6,289,342). Independent claims 1, 14, and 20 have been amended to further clarify the presently claimed invention, which distinguishes over Meyerzon taken alone and/or in view of Lawrence. Specifically, Meyerzon taken alone and/or in view of Lawrence is silent on:

retrieving a web document at an address, and extracting  
contents of the web document for rendering an intermediate  
dynamically constructed in-memory webpage representation of the

ARC9-2000-0046-US1

12

09/607,370

web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser;

loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein, wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage;

analyzing and summarizing the in-memory webpage representation to produce a text map for the webpage document of the textual contents; and

using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document.

Meyerzon discloses a web crawler program that includes a gatherer process for gathering information pertaining to electronic documents. See Meyerzon at col. 8, lines 58-60. Worker threads process URLs and then pass each URL to a filter daemon. See Meyerzon at col. 9, lines 13-16. The filter daemon uses the URL to retrieve the electronic document at the address specified by the URL. See Meyerzon at col. 9, lines 16-20. After retrieving an electronic, the filter daemon parses the electronic document and returns a list of text and properties. See Meyerzon at col. 9, lines 29-31. The worker thread then passes the list of properties and text to the indexing engine for creating an index which is used by the search engine in subsequent searches. See Meyerzon at col. 10, lines 13-16. A user may then examine the list of documents returned by the search engine, select a document and then, as stated by the Examiner on page 4 of the Office Action, the web browser displays the selected document to the user. See Meyerzon at col. 8, lines 23-25 and 32-35.

As can be seen from the diagram above, the presently claimed invention renders an in-memory representation of the webpage. Meyerzon does not teach, suggest, or

anticipate rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Meyerzon also does not teach, suggest, or anticipate loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein, wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage. Meyerzon renders the webpage at the client side and not before the hub processing unit analyzes and summarizes the in-memory webpage, as recited for independent claims 1, 14, and 20.

Further, Lawrence is completely silent on rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Meyerzon also does not teach, suggest, or anticipate loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein, wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage. Accordingly, independent claims 1, 14, and 20 of the present invention distinguish over both the Meyerzon and Lawrence references for at least this reason.

As the Examiner correctly states on page 7 of the Office Action, Meyerzon is silent on invoking an optical character recognition engine and analyzing an in-line GIF and IPEG images using the optical character recognition engine for text content. The Examiner goes on to combine Meyerzon with Lawrence.<sup>1</sup> The Examiner recites 35 U.S.C. §103.

---

<sup>1</sup> Applicants make no statement whether such combination is even proper.

The Statute expressly requires that obviousness or non-obviousness be determined for the claimed subject matter "as a whole," and the key to proper determination of the differences between the prior art and the present invention is giving full recognition to the invention "as a whole." The Meyerzon reference taken alone or in view of Lawrence simply does not suggest, teach or disclose the patentably distinct limitations of:

rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser;

loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein, wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage;

Continuing further, when there is no suggestion or teaching in the prior art for a hub processing unit for "extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser"; "loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation...wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage" the suggestion cannot come from the Applicants' own specification. The Federal Circuit has repeatedly warned against using the Applicant's disclosure as a blueprint to reconstruct the claimed invention out of isolated teachings of the prior art. See MPEP §2143 and Grain Processing Corp. v. American Maize-Products, 840 F.2d 902, 907, 5 USPQ2d 1788 1792 (Fed. Cir. 1988) and In re Fitch, 972 F.2d 160, 12 USPQ2d 1780, 1783-84 (Fed. Cir. 1992).

ARC9-2000-0046-US1

15

09/607,370



Moreover, the Federal Circuit has consistently held that when a §103 rejection is based upon a modification of a reference that destroys the intent, purpose or function of the invention disclosed in the reference, such a proposed modification is not proper and the *prima facie* case of obviousness cannot be properly made. See *In re Gordon*, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984).

Here the intent, purpose and function of Meyerzon taken alone or in view of Lawrence is the indexing of electronic documents for use by a search engine allowing a user to select a document to be displayed by a client-side web browser. The rendering of a webpage only occurs at the client side. Because Meyerzon does not render an in-memory representation of a webpage prior to analyzing and summarizing the in-memory webpage, this combination as suggested by the Examiner destroys the intent and purpose of Meyerzon. In contrast, the intent and purpose of the present invention is to render an in-memory webpage representation of a web document prior to analyzing and summarizing the in-memory webpage. Accordingly, the combination of Meyerzon and Lawrence results in an inoperable system, and the Examiner's case of "*Prima Facie Obviousness*" should be withdrawn.

Furthermore, the Federal Circuit stated in McGinley v. Franklin Sports, Inc., (Fed Cir 2001) that if references taken in combination would produce a "seemingly inoperative device," such references teach away from the combination and thus cannot serve as predicates for a *prima facie* case of obviousness. In re Sponnoble, 405 F.2d 578, 587, 160 USPQ 237, 244 (CCPA 1969) (references teach away from combination if combination produces seemingly inoperative device); see also In re Gordon, 733 F.2d 900, 902, 221 USPQ 1125, 1127 (Fed. Cir. 1984) (inoperable modification teaches away). Here, Meyerzon teaches rendering an electronic document for display on a web browser at the client side. Therefore, the combination of Meyerzon with Lawrence to produce the presently claimed invention where an in-memory webpage representation of a web document is rendered prior to analyzing and summarizing the in-memory

webpage would produce an inoperative device. Accordingly, the combination of Meyerzon and Lawrence is improper.

For the foregoing reasons, independent claims 1, 14, and 20 distinguish over Meyerzon taken alone or in view of Lawrence. Claims 2-13, and 15-19 depend from claims 1 and 14 respectively, either directly or by way of an intervening claim. Since dependent claims contain all the limitations of the independent claims, claims 2-13, and 15-19 distinguish over Meyerzon taken alone and/or in view of Lawrence, as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

Rejection under 35 U.S.C. §103(a) in view of Meyerzon, Lawrence with Hobbs

As noted above, the Examiner rejected claims 4-6 and 17-19 under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (U.S. Patent No. 6,638,314), in view of Lawrence et al. (U.S. Patent No. 6,289,342) as applied to claim 1 and in further view of Hobbs (U.S. Patent No. 6,523,022). With respect to Meyerzon and Lawrence, the above arguments regarding independent claims 1, 14, and 20 are applicable here and will not be repeated. As the Examiner correctly states on pages 8 and 9 of the Office Action, Meyerzon is silent on loading secondary documents including one or more Java applets with textual content embedded therein, loading secondary documents including web documents selected from the group of documents consisting of in-line frames, frames, and equivalents, and loading secondary documents including one or more Java Script components with textual content embedded therein, respectively. The Examiner goes on to combine Meyerzon with Lawrence as applied in claim 1, in further view of Hobbs.<sup>2</sup>

---

<sup>2</sup> Applicants make no statement whether such combination is even proper.

Hobbs discloses dynamically augmenting the contents of an information file on a first network resource. See Hobbs at col. 7, lines 53-55. Hobbs is completely silent on rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Hobbs also does not teach, suggest, or anticipate loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein, wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage. Accordingly, independent claims 1, 14, and 20 of the present invention distinguish over Meyerzon and/or Lawrence and/or in further view of Hobbs for at least this reason.

Continuing further, when there is no suggestion or teaching in the prior art for a hub processing unit for "extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser"; "loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation...wherein the hub processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage" the suggestion cannot come from the Applicants' own specification. The Federal Circuit has repeatedly warned against using the Applicant's disclosure as a blueprint to reconstruct the claimed invention out of isolated teachings of the prior art. See MPEP §2143 and Grain Processing Corp. v. American Maize-Products, 840 F.2d 902, 907, 5 USPQ2d 1788 1792 (Fed. Cir. 1988) and In re Fitch, 972 F.2d 160, 12 USPQ2d 1780, 1783-84 (Fed. Cir. 1992).

Moreover, the Federal Circuit has consistently held that when a §103 rejection is based upon a modification of a reference that destroys the intent, purpose or function of the

invention disclosed in the reference, such a proposed modification is not proper and the *prima facie* case of obviousness can not be properly made. See *In re Gordon*, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984).

Here the intent, purpose and function of Meyerzon taken alone or in view of Lawrence and/or in further view of Hobbs is the indexing of electronic documents for use by a search engine allowing a user to select a document to be displayed by a client-side web browser. The rendering of a webpage only occurs at the client side. Because Meyerzon does not render an in-memory representation of a webpage prior to analyzing and summarizing the in-memory webpage, this combination as suggested by the Examiner destroys the intent and purpose of Meyerzon. In contrast, the intent and purpose of the present invention is to render an in-memory webpage representation of a web document prior to analyzing and summarizing the in-memory webpage. Accordingly, the combination of Meyerzon and Lawrence in further view of Hobbs results in an inoperable system, and the Examiner's case of "*Prima Facie Obviousness*" should be withdrawn.

Furthermore, the Federal Circuit stated in McGinley v. Franklin Sports, Inc., (Fed Cir 2001) that if references taken in combination would produce a "seemingly inoperative device," such references teach away from the combination and thus cannot serve as predicates for a *prima facie* case of obviousness. In re Sponnoble, 405 F.2d 578, 587, 160 USPQ 237, 244 (CCPA 1969) (references teach away from combination if combination produces seemingly inoperative device); see also In re Gordon, 733 F.2d 900, 902, 221 USPQ 1125, 1127 (Fed. Cir. 1984) (Inoperable modification teaches away). Here, Meyerzon teaches rendering an electronic document for display on a web browser at the client side. Therefore, the combination of Meyerzon with Lawrence and/or in further view of Hobbs to produce the presently claimed invention where an in-memory webpage representation of a web document is rendered prior to analyzing and summarizing the in-memory webpage would produce an inoperative device. Accordingly, the combination of Meyerzon and Lawrence in further view of Hobs is

improper.

For the foregoing reasons, independent claims 1, 14, and 20 distinguish over Meyerzon and/or Lawrence and/or in further view of Hobbs. Claims 4-6 and 17-19 depend from claims 1 and 14 respectively either directly or by way of an intervening claim. Since dependent claims contain all the limitations of the independent claims, claims 4-6 and 17-19 distinguish over Meyerzon and/or Lawrence and/or in further view of for Hobbs as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

### CONCLUSIONS

The remaining cited references have been reviewed and are not believed to affect the patentability of the claims as previously amended.

In light of the Office Action, Applicants believe these amendments serve a useful clarification purpose, and are desirable for clarification purposes, independent of patentability. Accordingly, Applicants respectfully submit that the claim amendments do not limit the range of any permissible equivalents.

Applicants acknowledge the continuing duty of candor and good faith to the disclosure of information known to be material to the examination of this application. In accordance with 37 CFR §§ 1.56, all such information is dutifully made of record. The foreseeable equivalents of any territory surrendered by amendment is limited to the territory taught by the information of record. No other territory afforded by the doctrine of equivalents is knowingly surrendered and everything else is unforeseeable at the time of this amendment by the Applicants and their attorneys.

Applicants respectfully submit that all of the grounds for rejection stated in the Examiner's Office Action have been overcome, and that all claims in the application are

ARC9-2000-0046-US1

20


09/607,370

allowable. No new matter has been added. It is believed that the application is now in condition for allowance, which allowance is respectfully requested.

**PLEASE**, if for any reason the Examiner finds the application other than in condition for allowance, the Examiner is invited to call either of the undersigned attorneys at (561) 989-9811 should the Examiner believe a telephone interview would advance the prosecution of the application.

Respectfully submitted,

Date: April 27, 2005

By:   
Jon Gibbons  
Registration No. 37,333  
Attorney for Applicants

FLEIT, KAIN, GIBBONS,  
GUTMAN, BONGINI & BIANCO P.L.  
One Boca Commerce Center, Suite 111  
551 Northwest 77th Street  
Boca Raton, FL 33487  
Tel. (561) 989-9811  
Fax (561) 989-9812

Please Direct All Future Correspondence to Customer Number **23334**

ARC9-2000-0046-US1

21

09/607,370